IMPLICIT FEEDBACK: A NEW WAY OF MEASURING PATIENTS INTERESTS IN WEB-BASED HEALTH LITERATURE

By

Zhou Yan

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Biomedical Informatics

May 2013

Nashville, Tennessee

Approved:

Professor Samuel Trent Rosenbloom

Professor Gretchen Purcell Jackson

Professor Cynthia S Gadd

Professor Qingxia Chen

To John

ACKNOWLEDGEMENTS

I am grateful to all my committee members for their support and guidance throughout this journey.

I would also like to acknowledge the iADAPT team for their broad expertise and continuing support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES AND EQUATIONS

CHAPTER I

INTRODUCTION

With growing access to the Internet and online health information, more people are looking online for answers to their health related questions. According to a report published by PEW International Research for the year 2012, one in three American adults have gone online to self-diagnose a medical condition. Among Internet users in the United States (U.S.), 72% have searched online for healthcare- related information. [1] Meanwhile, U.S. healthcare organizations are putting more emphasis on personalized and patient-centric care. [2] Healthcare organizations, in their effort to provide quality service for patients who seek information online, are demanding research measuring patients' needs and interests in web-based healthcare literature. [3]

There are two traditional approaches to evaluating individuals' interest in an information resource. The first approach is conducting cognitive research on volunteer participants in an experimental setting. The strength of this approach is that it produces an in-depth understanding of participants' attitude and interests. However, this approach is often associated with high financial and time costs. [4] The second approach is user surveys. [5] Survey administrators seek participants' opinions through questionnaires, interviews or focus group meetings. User surveys generally have lower costs in comparison to cognitive research, but response rates have been declining since the 1990s. [6,26] Participants often skip surveys when there is not a direct benefit or the survey requires too much time. [52]

To address the problems associated with these existing methods, the current research project proposes an implicit feedback approach. Instead of explicitly asking the users' opinions, this implicit feedback approach observes the user's behaviors as they interact with an information resource. The behavior-oriented data is employed to infer users' interests.

The implicit feedback approach has several advantages. First, it does not require a laboratory setting or dedicated user effort, both of which have monetary and time costs. Data collection occurs automatically and unobtrusively while users browse an information resource in their normal environment. The implicit feedback approach also captures data from every user who interacts with the information resource. This complete user coverage addresses the problem of the low response rates that bias the results of user surveys.

This masters' thesis describes an evaluation of implicit feedback that leveraged an existing Diabetes Education page in a widely adopted patient portal. The study collected implicit feedback from 1450 users of this information resource over a six-month period. Two types of implicit feedback were assessed: Page Staying Time and Link Count. Page Staying Time measured the total amount of time user stayed on a web page. The Link Count determined the number of hyperlinks followed by the user. An overview of this research project is outlined in four steps below.

Step 1: Develop a Tool to Capture Implicit Feedback. This research project developed an extensible tracking tool to capture implicit feedback unobtrusively, based on users' interaction with a Diabetes Education web page embedded in a patient portal.

Step 2: Measure Implicit Feedback. Algorithms designed as part of this study extracted implicit Page Staying Time and Link Count measures.

Step 3: Establish a Reference Standard. Online survey tools added to the Diabetes Education page as part of this study allowed users to indicate whether information on the web page was useful. Voting behavior served as a surrogate of patient users' interests and an explicit feedback reference standard to evaluate the implicit measures.

Step 4: Explore the Relationships Between Explicit and Implicit Feedback Measures. The research study team built regression models to explore the relationships between the explicit voting behaviors and implicit Page Staying Time and Link Count measures.

To summarize, the research study described in this thesis makes three informatics contributions: 1) an extensible tracking tool to collect automatically and unobtrusively users' interactions with web pages 2) a partial time algorithm to calculate the Page Staying Time 3) an evaluation of implicit feedback to assess users' interest in web-based health literature.

CHAPTER II

BACKGROUND AND SIGNIFICANCE

As the amount of online health-related literature has increased, so has its impact on patient education. [7] Evaluation methods to assess interest in the online health literature are critical to creating effective health information resources that meet information needs. [8]

Existing Solutions

User surveys have been considered one of the gold standards for assessing attitudes, beliefs and interests. [9] Questionnaires elicit participants' opinions and ratings. User surveys have proven strengths in providing an in-depth understanding of users' attitudes and interests. [24] However, declining response rates challenge the utility of user surveys. [6,26] The American Association for Public Opinion research conducted a study on survey response. The response rates across all modes of survey declined over the last decade. [11,12,13] As a result, organizations that rely on user surveys must devote more effort to survey administration. One practice is to provide incentives to the study participants. [14,15,16] This practice increases the cost of a survey and raises concerns about the influence of incentives on the accuracy on the survey results. [17,18,19,20] In addition to financial costs, low response rate increase the risk of potential non-response bias. [21]

Cognitive studies, as applied in usability tests, are another established method frequently used to evaluate users' reactions and interests. [22,23] In cognitive studies, participants are invited to a dedicated laboratory environment to interact with an information resource. Reactions to a resource are carefully observed and recorded for further analysis. The high financial costs, limited number of available laboratories dedicated to this kind of research and time required for conducting cognitive studies limits their applications.

Confined by these limitations, user interest studies have often been conducted with a limited number of participants in a laboratory setting, instead of with a large population in a real-world setting. [43, 44, 45] In the case of the Internet health information resources, the relevant study domain includes the worldwide population of healthcare consumers. A new solution is needed to fill in the research gap in evaluating the interests' of these users.

Alternative Approach

To address the challenges of existing solutions, this thesis reports a study that evaluated an alternative approach to evaluating users' interests: collecting and analyzing implicit feedback. *Implicit feedback* is also known as *implicit measure* or *implicit rating* in different literatures. Leverage existing research, this study defines implicit feedback as any data that can be collected unobtrusively by observing a user's interaction with an information resource. [27-29, 46-50] Instead of explicitly soliciting opinions, the implicit feedback approach observes users' behaviors while they interact with an information resource. The behavior-oriented data can be analyzed to derive users' opinions and interests. Contrary to the cognitive approach, behaviors and reactions are observed in a natural setting instead of a laboratory environment.

The implicit feedback concept is not new. Research on implicit feedback originated in the Information Retrieval (IR) field. Information resources under study could be web documents, Internet news articles, movies, or television programs. [25] To develop and improve IR systems, researchers conducted Cranfield-style evaluations to measure the precision (i.e., percentage of retrieved items that are relevant) and recall (i.e., the percentage of relevant articles that are retrieved). [46] A major challenge to these types of evaluations was the measurement of "relevance". Cranfield-style evaluations conduct the relevance assessment by involving domain experts to assign a relevance level to each retrieval results. To make this approach practical, most experiments were done in small collections of test documents, for which relevance was determined in a time-consuming and labor-intensive manual process. The measurement of relevance in a laboratory setting is highly suspect, as the relevance of an information resource is often context dependent. To solve these problems, researchers looked into measurable user interactions with the information resource, reflected in various forms of implicit feedback. The implicit feedback was then used as a measure of relevance to determine precision and recall, which were optimized to augment information retrieval, filtering and recommendation. [60]

Nicoles in his 1997 study evaluated the costs and benefits of using implicit measures to replace the explicit ratings. His study looked at thirteen potential types of implicit actions including read time, save/print action, and marking document as favorite. The study results suggested that implicit ratings have a good potential for being able to predict user satisfaction without interrupting users' normal workflow. [46,47] Morita & Shinoda studied how much time users spent on Usenet news articles and concluded that reading time could predict users' interest level. [48] Konstan et al.'s 1997 study further confirmed Morita & Shinoda's finding that reading time was a strong predictor of user interest. [49] Oard & Kim's study further broadened the concept of implicit measure to include users' behavior in information retention as in the case of printing a page. Their study found that reading time as well as whether a user prints a page was a useful indicator of user interest. [50] Kelly & Teevan's 2003 study provided three important conclusions on implicit measures: "First, there is good potential for implicit measures to either replace or act in conjunction with explicit ratings or feedback. Second, there is some disagreement in the existing research on *exactly* which implicit measures are useful – at least within the domain of search engines. Finally, most of the studies have been conducted in laboratory settings." [51]

Search Engine Optimization (SEO) is another domain in which investigators have conducted extensive research using relevant implicit feedback with a goal of reducing the dependence on explicit human judgments. [27] With SEO, machine-learning algorithms are developed using implicit feedback to improve retrieval quality. Examples of implicit feedback that have been studied include the links a user clicks on in the ranked search results, the time a user spends reading a resulting page, or how a user reformulates a query. [28,29,30]

To summarize, measures of implicit feedback that have been investigated include reading time, page scrolling, printing, and click-through. [28,30,46–50,54] Study results have shown strong correlation between implicit measures and user interest and satisfaction. However, these studies have largely been conducted in controlled laboratory settings with customized web browsers and limited numbers of users. The extent to which existing research applies to real-world settings is unclear. [52-60]

This thesis proposes that collection and analysis of implicit feedback may be a promising alternative approach to evaluating users' interests in web-based information by leveraging previous research on the use of implicit feedback in the IR and SEO domains.

The hypothesis behind the research presented in this thesis is that implicit feedback, based upon users' interactions with a web-based information resource, is correlated with users' interests in the information resource, as measured by explicit feedback. To test this hypothesis, we measure the use of an existing Diabetes Education page hosted inside a patient-facing portal web site – My Health At Vanderbilt (MHaV).

The Diabetes Education page (iADAPT) was developed as a component of a research project funded by the Agency for Healthcare Research and Quality (AHRQ) through its Innovative Adaptation and Dissemination of AHRQ Comparative Effectiveness Research Products (iADAPT) program (Grant No. R18 HS019276, Principal Investigator: Samuel Trent Rosenbloom). The iADAPT project's aim is to transform the AHRQ's evidence-based diabetes guidelines into a concise, targeted and easily actionable information resource for patients and healthcare providers. This resource is then disseminated to patients with Type II diabetes through a widely used patient portal – My Health at Vanderbilt (MHaV). [31,68]

MHaV, developed at Vanderbilt University Medical Center in 2002 provides Vanderbilt patients with free, online access to portions of their electronic medical records. MHaV also allows patients to check and schedule appointments, refill prescription, pay bills, view test results, immunization records and a list of their medications. Other popular features provided by MHaV include secure messaging between patients and their healthcare providers and personalized health information based on the patients' diagnosis. [32]

This thesis project leverages the Diabetes Education information resource placed into MHAV as part of the iADAPT program. All collection and analysis of implicit feedback was conducted on the iADAPT Diabetes Education page. Approval for tracking patients' usage of this page was approved by the Vanderbilt Institutional Review Board (IRB) for the iADAPT project.

Significance

The implicit feedback measured in this project possesses several unique characteristics. This section lists five major characteristics of the implicit feedback, along with a statement about the significance of each in serving as an attractive solution in evaluating users' interests.

1) Complete user coverage

This study collects implicit feedback data from every user's interaction with the Diabetes Education page. Consequently, the study covers the complete user base and eliminates the low response rate problem often associated with user surveys.

2) Richness of data

The implicit feedback approach captures data for every user during every visit to the page. Depending on the contents of the web page, user interactions can be captured at a very granular level. Examples of potential user interactions could include: button clicks, link clicks, mouse movements, scroll bar movements, movie or sound track playing, and keyboard typing.

3) Low Cost

The implicit feedback approach removes the cost associated with survey administration and laboratory settings required by cognitive studies. With the help of widely available high-throughput computer servers, large amounts of data can be quickly retrieved at a low cost. Advanced data mining and Online Analytic Processing (OLAP) can further facilitate and lower the cost of data processing and analyzing.

4) Avoiding self-selection bias

Self-selection bias, often observed in user surveys, is caused by the fact that certain groups of users are more likely to provide feedback than other groups. For instance, users who are extremely happy or extremely unhappy about an application are more likely to offer their opinions, compared to the user who has a neutral view. This unbalanced sample selection will influence survey results by over-representing groups with certain characteristics and thus distorting results. [33] Implicit feedback, due to its all-inclusive and non-disruptive nature, covers the complete range of the user base, thus effectively avoiding self-selection bias.

5) Potential in trend prediction

User surveys, often conducted discretely due to their high cost, present a point-in-time and limited-context assessment of the users' interest. Implicit feedback, on the contrary, is dynamic in nature. It can be captured in actual user settings and stored for as long as needed with a relatively low cost. Statistical models built on top of the continuous implicit feedback data represent a user's view over time, therefore providing the potential to detect user behavior patterns and predict future trends. This type of assessment is an important contribution because user interest is expected to be context-dependent and dynamic variable.

CHAPTER III

METHODS

This study leveraged the existing iADAPT project to evaluate users interest in the Diabetes Education page developed by the iADAPT team. The design and layout of the study is outlined in four steps: Step1) Develop an extensible tracking tool to capture implicit feedback while the iADAPT study participant interacts with the iADAPT page. Step 2) Apply an algorithm to quantify the collected implicit feedback and generate two types of implicit measures. Step 3) Establish a reference standard to evaluate users interests. Step 4) Build regression models to analyze the relationship between the implicit feedback and the explicit feedback reference standard.

Study Setting

This evaluation of implicit feedback will be conducted on a Diabetes Education page that is deployed in the widely-adopted Vanderbilt patient portal, MyHealthAtVanderbilt.com (MHaV). MHaV is a patient portal web-based tool that allows Vanderbilt patients and their families to interact with the healthcare system. Developed in 2002 at the Vanderbilt University Medical Center (VUMC), MHaV had over 220,000 registered users at the end 2012, which represents over 25% of all potential users. MHaV has a very active user base with an average over 11,000 visits per day. [69-71]

As an integral component of the iADAPT project, the Diabetes Education page included diabetes-related health information. Three content-specific sections composed this page: 1) Personalized data including the patient's current prescription medicines and common drug side effects, 2) Authoritative literature such as the AHRQ diabetes guideline for patients and physicians about the medicines for diabetes, and 3) Recommended readings about physical activity, nutrition information, and general diabetes knowledge. This iADAPT page was exposed to qualified iADAPT study participants during a six-month study period.

Qualification Criteria: The study population for this research projected included participants in a larger iADAPT project to disseminate diabetes evidence through a patient portal. The eligibility criteria for patients in this research project were Vanderbilt patients with: 1) Age 18 years or older, 2) Registration for and use of a MHaV account at least twice during the year after registration, and 3) an International Classification of Diseases (ICD 9) diagnosis code for Type II Diabetes.

Capture Implicit Feedback

This research project developed a tracking tool designed unobtrusively to capture patients' interaction with the various components of the MHaV Diabetes Education page and to capture implicit feedback from within the page. This page was made available to all patients who met eligibility criteria within MHaV. The MHaV Diabetes Education page is shown in Figures 1-5, and a red circle identifies each tracked component. Due to the vertical length of the Diabetes Education page when viewed on the computer screen, the contents are shown in five sections here in the thesis. The first section is Diabetes

Medicines (Figure 1). This section shows a list of the patient's prescribed diabetes medicines. A multimedia play button on the right side of each medicine plays an audio file that pronounces the medication's name upon selection.

The patient can select the "If you have any questions about your diabetes or medications" link to communicate with his or her provider regarding the prescribed medicine. This communication is conducted through a secure messaging system and is automatically recorded in the patient's electronic medical record (EMR) for reference. [75]

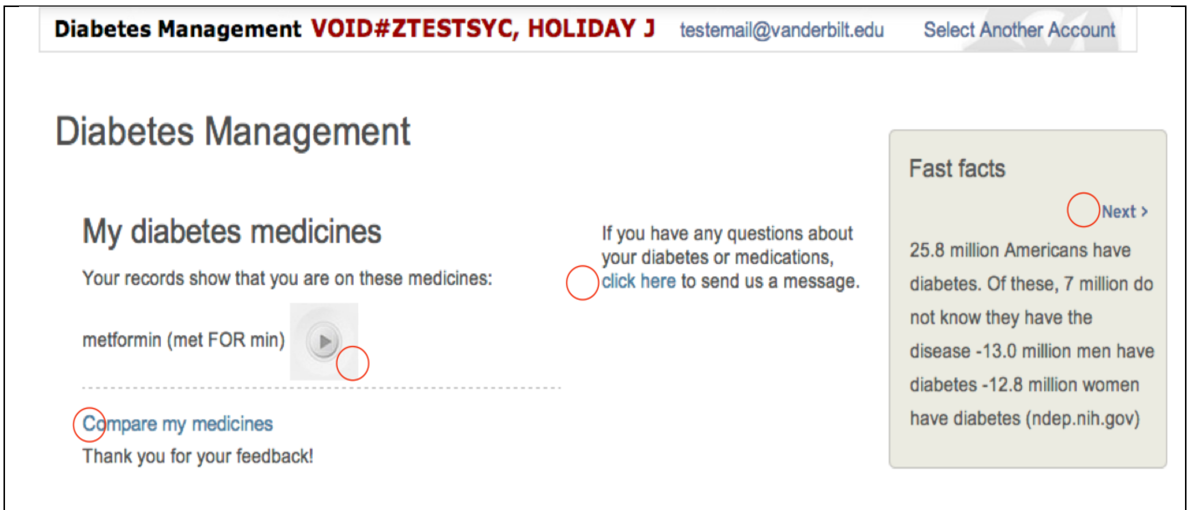

Figure 1 Diabetes Medicine Section in the Diabetes Education Page

The link "Compare my medicines" displays a drug comparison table cross-referenced with the common side effects (Figure 2).

Side Effects

Comparing your medicine note: All type II diabetes medicines below can help with blood sugar. Sometimes, these medicines can make your blood sugar too low. Test your blood sugar to make sure you keep good blood sugar levels.

Medicine Name	weight gain	low red blood cells	acid in bloodstream	heart failure	bad cholesterol	bad fat	stomach problems	liver enzymes	swelling	immune system	bone fractures
Acarbose	☺					☺	☹				
Actoplusmet	☹	☹	☹	☹	☹	☺		☹	☹	☹	☹
Avandamet	☹	☹	☹	☹	☹			☹	☹	☹	☹
Avandaryl	☹	☹		☹		☹		☹	☹	☹	☹
Chlorpropamide							☹				
Glyburide	☹					☺	☹	☹			
Glimepiride	☹					☺		☹			
Glipizide	☹					☺	☹	☹			
Glucovance	☹		☹		☺	☺		☹			
Janumet			☹			☺		☹			
Metaglip	☹		☹		☺			☹			
Metformin 1	☺		☹		☺	☺	☹	☹			
Miglitoln 2	☺						☹				
Nateglinide										☹	
Pioglitazone 3	☹	☹		☹	☺☹	☺		☹	☹	☹	☹
Repaglinide	☹					☺					
Rosiglitazone	☹	☹		☹	☺☹	☹		☹	☹	☹	☹
Sitagliptin											
Tolazamide							☹				
Tolbutamide							☹				

*We used Clinical Pharmacology, Micromedex, Facts and Comparisons, and the AHRQ literature to compare the most common side effects based on percentages from clinical trial data and those that affect the patients quality of life the most (weight gain). Side effects and benefits were determined for each drug and not the drug's class.

1- This medicine may cause heart disease or death. 2- This medicine may cause a skin rash. 3- This medicine may cause kidney failure.

Figure 2 Drug Comparison Table in the Diabetes Education Page

The second section is Diabetes Literature. As shown in Figure 3, this section lists four diabetes management guidelines. Medicines for Type 2 Diabetes, Doctor's Guide to Type 2 Medicines and Type 2 Diabetes Research Summary are provided and published by the AHRQ. Quick Type 2 Diabetes Summary Guide is an adapted version of the AHRQ literature created by the iADAPT project team at Vanderbilt University.

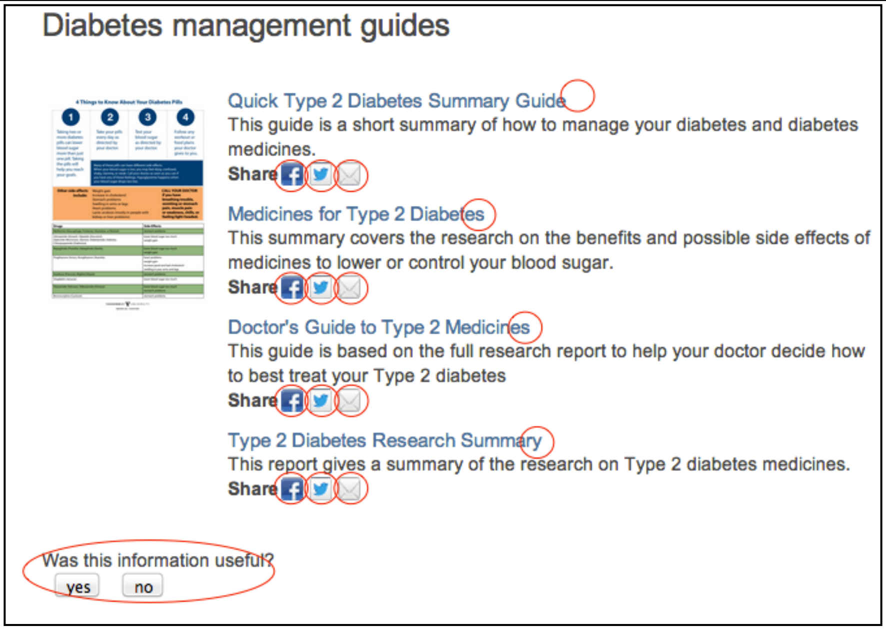

Figure 3 Diabetes Literature Section in the Diabetes Education Page

Each guideline contains a brief description of the contents. A cover image of the literature is displayed when a user's mouse hovers over the link. Shareable social media links including Facebook, Twitter, and email are listed below each guideline.

The third section is the Recommended Readings for diabetes patients. The readings are categorized into three domains (i.e., Physical Activity, Nutrition, and About Diabetes), each represented by a tabbed page.

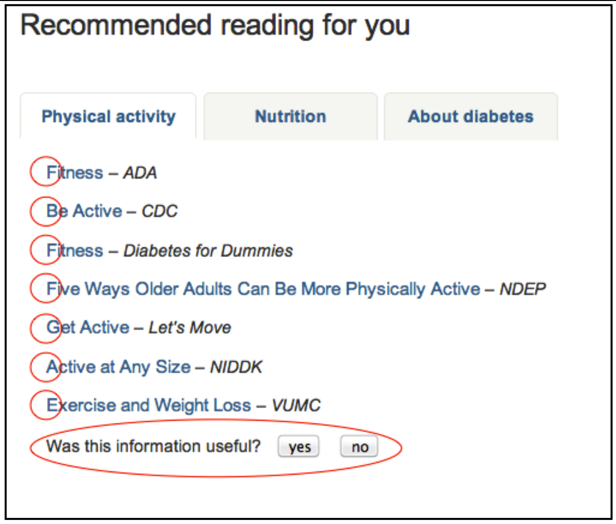

Figure 4 Recommended Reading – Physical Activity

Figure 4**Error! Reference source not found.** shows the Physical Activity Page. It lists seven recommended readings related to the exercise information for the diabetes patients.

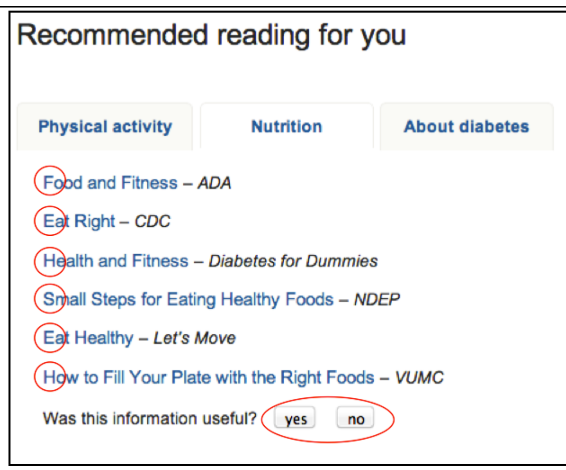

Figure 5 Recommended Reading – Nutrition

Figure 5shows the Nutrition Page. A group of six recommended readings direct the user to nutritional information appropriate to diabetes patients.

Figure 6 shows the About Diabetes Page. This page lists general information regarding diabetes disease and a graphic illustration of the relationships among food, sugar, and insulin.

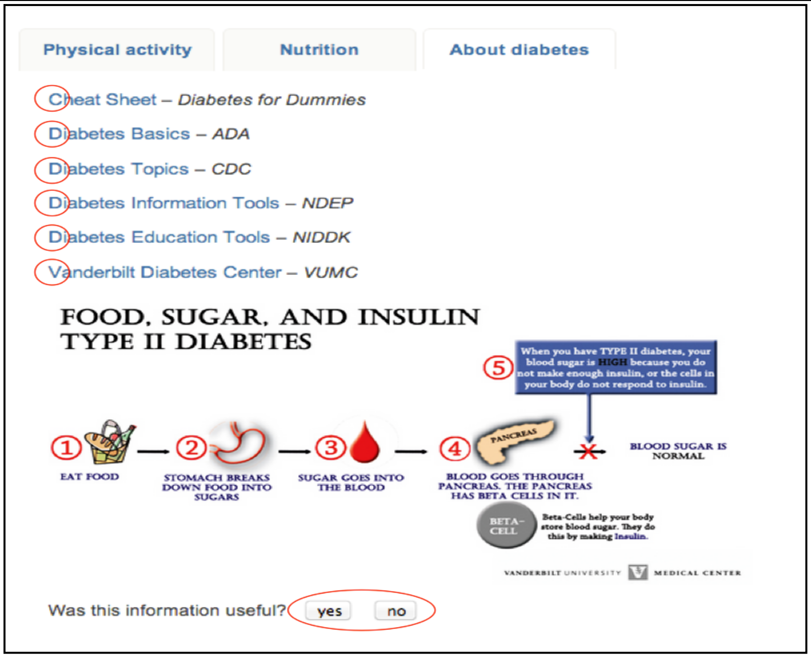

Figure 6 Recommended Reading – About Diabetes

While a patient browsed the Diabetes Education page, his or her interaction with the web page was logged as events. *Events* are defined as actions a user performs on a web page. For example, when a user clicks a button or link, moves the mouse over an element, types in a text field with the keyboard, or scrolls up and down the page with a scrolling bar, each of the above actions generates an individual event. In this project, an extensible tracking tool captured the user-web page interaction events.

The tracking tool functions are illustrated in a flow chart in Figure 7, which illustrates what happens when a patient clicks on the Sound Play Button next to the medicine “Metformin”. This patient-web page interaction in turn registers a button click event from the patient’s browser.

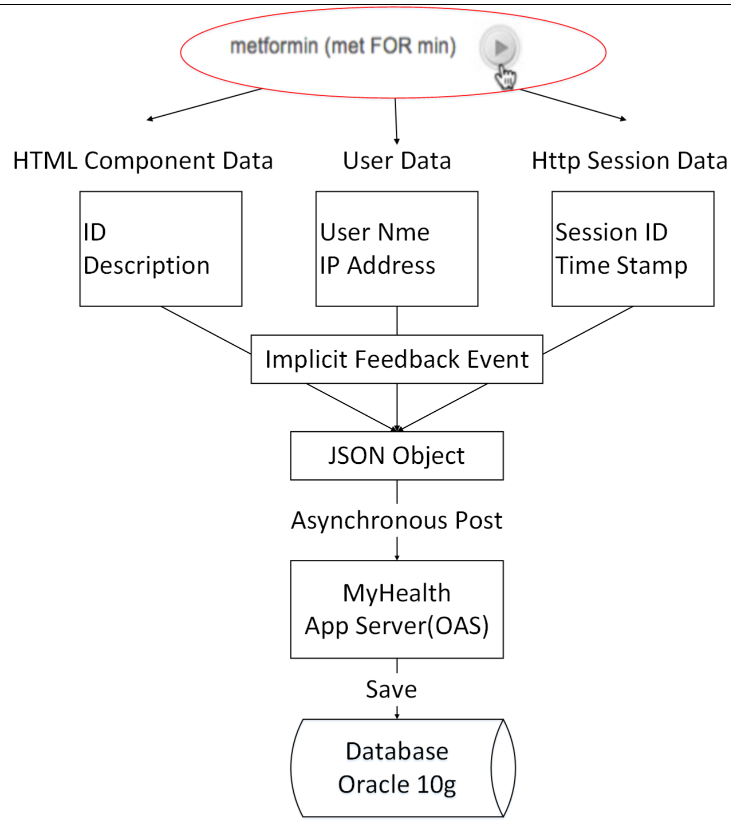

Figure 7 Implicit Feedback Tracking Tool in Flow Chart

The tracking tool was designed to react to the sound button click event in four chronological steps as summarized below.

Step 1: The details of the event were collected in three components: 1) HTML [34] component data including the component’s id and description, 2) User data including the patient’s MHaV user name and the Internet Protocol (IP) address [35] of the user’s computer, 3) Session data including the session ID and the timestamp of the event occurring time.

Step 2: The tracking tool packages detailed event data to form an implicit feedback event. This implicit feedback event was formatted into a JavaScript Object Notation (JSON) object. [36, 37] This tool used JSON objects to exchange data among applications for several reasons. JSON is a lightweight XML; it is language independent and is supported by a broad range of software libraries. [38]

Step 3: The tracking tool posted the JSON object to the backend OAS server that hosts MHaV. The tracking tool conducted the posting through Asynchronous JavaScript and XML (AJAX) calls. [39] This study used the AJAX approach due to its asynchronous communication nature. [76] With AJAX, communication between the client browser and the server occurs in the background without interrupting user's natural browsing action. [40]

Step 4: After the MHaV server received the JSON object, the server saved the details of the implicit feedback data to a table in the Oracle 10g database. [41, 42]

This tracking tool, developed for this research, was designed to be extended to a broad range of web pages that support HTML and JavaScript.

Measure Implicit Feedback

During a six-month study period, the tracking tool continuously captured, processed, and saved implicit feedback events as the study participants interacted with the MHaV Diabetes Education page. Table 1 illustrates examples of the implicit feedback as represented by user interactions saved in the database. Also, the Diabetes Education page-specific events are mixed with other MHaV portal activities.

Table 1 Sample of Study Collected Implicit Feedback Data

ID	User ID	Event Time	Event Name	Description
1	Steve	09:00	Login	Steve logged in to MHaV successfully
2	Steve	09:01	View Diabetes Education Page	Steve opened the Diabetes Education Page
3	Steve	09:02	Sound Play	Steve clicked the medicine pronunciation button
4	Mike	09:03	Login	Mike logged in to MHaV successfully
5	Steve	09:04	Diabetes_meds_yes	Steve voted yes on the Diabetes Medicine section
6	Mike	09:05	Ahrq_pdf	Mike clicked an AHRQ literature link
7	Mike	09:06	Ahrq_pdf_yes	Mike voted yes on the AHRQ literature section
8	Steve	09:07	Logout	Steve logged out
9	Mike	09:10	View Lab Page	Mike opened the Labs Page

To measure implicit feedback, an algorithm was developed to take the implicit feedback as input and produce implicit measures for each study participant. Appendix A shows a conceptual flow chart of how this algorithm works. This algorithm first filtered events that are not associated with the Diabetes Education page. Next, the algorithm grouped events by patients. The algorithm then iterated through each event for each study participant and passed the activity through a series of decision points as shown in the flow chart. Outcomes from the decision points were tallied to produce the Page Staying Time and link count implicit measures. When the program reaches the last activity of each study participant, the script generated a summary of all implicit measures for a user. The program then repeated the steps above until all study participants' data have been processed.

The resulting implicit measures were measured as the sum of the Page Staying Time and Link Counts during a participants' six-month interaction with the MHaV Diabetes Education page (Table 2).

Table 2 Implicit Measure Types and Descriptions

Type	Implicit Measure	Description of the Implicit Measure
Time	Page Staying Time (minutes)	The total amount of time that a user stayed on the Diabetes Education Page
Link Count	Sound Play	The total number of times the user clicked on the medicine's Sound Play Button
	AHRQ Share	The total number of times the user clicked on the Facebook, Twitter or email links beneath each of the four AHRQ literatures
	AHRQ Link	The total number of times the user clicked on the four AHRQ literature links
	Drug Table	The total number of times the user clicked on the "compare my medicines" link
	Facts	The total number of times the user clicked on the "Prev" or "Next" button in the Fast Facts
	Activity Link	The total number of times the user clicked on the links inside the Physical Activity page
	Nutrition Link	The total number of times the user clicked on the links inside the Nutrition page
	About Link	The total number of times the user clicked on the links inside the About Diabetes page
	Messaging	The total number of times the user clicked on the "send us a message" link

The resulting Link Counts consists of 9 individual link measures. Each link measure represents the total number of times the user clicks on a particular hyperlink. The calculation of the link measure is straightforward: every mouse click on a hyperlink was increased the link measure by 1.

The calculation of the Page Staying Time was more involved due to the complex nature of web browsing. In a simple browsing scenario (as illustrated in Figure 8) a user would open the Diabetes Education page at 14:01 and later switch to a different web page or logs out of the web site at time 14:02. The Page Staying Time for the Diabetes Education page was calculated as the time between 14:01 and 14:02. In this example, the resulting Page Staying Time is 1 minute.

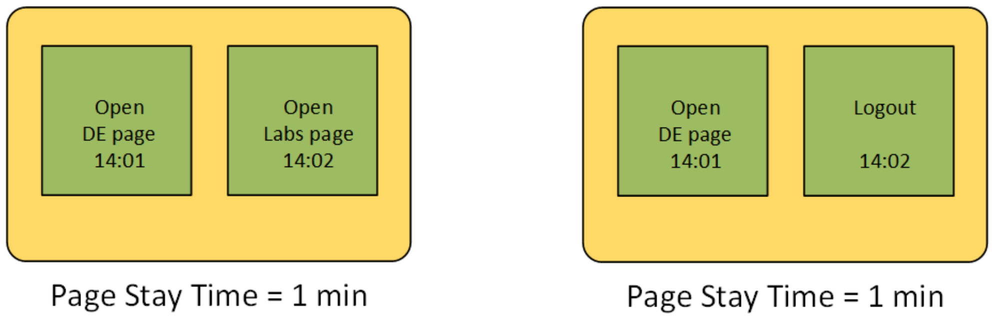

Figure 8 Page Staying Time – Scenario 1

In a more complex scenario, a user would open a web page and then get distracted or move on to other activities without either logging out or switching to a different page. In this case, the opened page will be left in an abandoned state until the user's web session expires. In the case of MHaV, web sessions expire after no user activity is detected in 30 minutes. Figure 9 illustrates the abandoned page scenario.

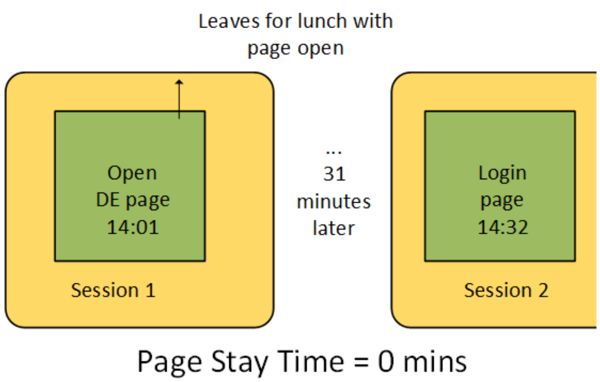

Figure 9 Page Staying Time – Scenario 2

To handle time calculation for the abandoned pages, a common practice, as practiced by Google Analytics, is to disregard these data.

> "*When a page is the last page in a session, there is no way to calculate the time spent on it because there is no subsequent pageview. For this reason, when Page A is the last page in the visitor's session, its time calculation is not counted for that view. In addition, when that page is the only page viewed in the session, no time on page is calculated.*" [63]

Google's approach is simple yet neglects useful information. The fact that a page is the last one in the session does not necessary mean the user spent zero amount of time on the page. To solve the problem, the research team developed a partial-time accumulation algorithm. With this algorithm, when a user opens the Diabetes Education page, an internal timer marks this timestamp as the starting point. As the user conducts further activities on this page, such as clicking links, clicking buttons or casting votes, this new activity's timestamp will be preserved as a temporary ending point. As the user continues his or her activity on a given page, the temporary end point gets updated with each new activity's timestamp. Figure 10 illustrates how the partial-time algorithm calculates the page staying time of an abandoned page with two user events.

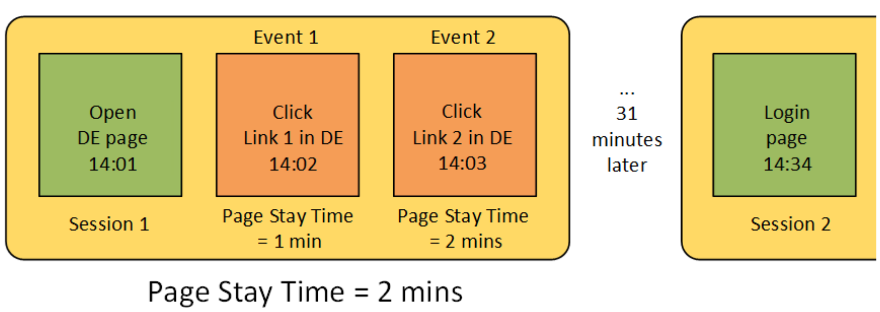

Figure 10 Page Staying Time – Scenario 3

In the case where the user switches to another page or logs out, the end point marks the completion of the user's stay on the page. In the situation where the user abandons the page, the last temporary ending point, which indicates the user's last activity on the page, marks the end of the partial page time accumulation. This measure is a conservative estimate of time spent on a page because users may continue to read information even after the last measurable activity is executed.

Establishing a Reference Standard

To evaluate patients' interest with an explicit feedback metric, the research team built online survey tools inside the MHaV Diabetes Education page to collect participating patients' votes about perceived usefulness. Each survey tool asked: "Was this information useful?" and provided "yes" and "no" voting buttons. This survey questions were placed at five locations in the Diabetes Education page (shown in Figures 1, 3-6).

Users had the option to vote "yes", vote "no", or ignore the questions by not voting. Once a vote was cast, the voting button was replaced by a message stating, "Thank you for your feedback." Users were not able to vote on the same question more than once in the same visit. When the user returned to the page on a subsequent login, all voting questions became available. In the case of a user casting multiple votes on the same question, the last vote was considered to override any previous votes.

Table 3 summarizes the intended meaning of the five survey questions.

Table 3 Explicit Feedbacks Based on Patient's Online Vote

Explicit Feedback	Value	Description
MEDS_VOTE	"yes", "no" or n/a	User's vote on usefulness of the Diabetes Medicines section
AHRQ_VOTE	"yes", "no" or n/a	User's vote on the usefulness of the AHRQ literature section
ACTIVITY_VOTE	"yes", "no" or n/a	User's vote on the usefulness of the Physical activity section
NUTRITION_VOTE	"yes", "no" or n/a	User's vote on the usefulness of the nutrition section
ABOUT_VOTE	"yes", "no" or n/a	User's vote on the usefulness of the general diabetes section

The analysis of these data dichotomizes the voting activity into "Voted" and "Not Voted" groups. The Voted group represents study participants who casted at least one vote during the study period. The Not Voted group includes study participants who accessed the Diabetes Education page but never cast any vote during the study period. The current study applied this binary voting result as a surrogate for evaluation of patients' interest. The "Voted" behavior implies patients' interest and the "Not Voted" behavior infers patients' lack of interests in the Diabetes Education page.

This study used patients' voting behavior as a surrogate for patients' interests. This approach was based on prior research studying the correlation between survey participation rate and topic interest. When the content topic is of interest to an individual, that person is more likely to participate in the survey by casting votes. Groves et al. found that "persons cooperated at higher rates to surveys on topics of likely interest to them. The odds of cooperating are roughly 40 percent higher for topics of likely interest than for other topics." [64] Holland and Christian's study on the influence of topic interests also found high nonresponse rates for people who are less interested in the topic. [65] It should be noted that a user's actual interest is inherently a continuous variable. The decision to simplify the user's interest into a binary outcome is one limitation of this study.

Nonparametric Analysis

Initial results from this study indicated that the implicit measure do not follow a normal distribution. As we cannot make assumptions about normality, parametric analysis was not applicable to this project. As a result, we conducted nonparametric analysis on the implicit measures between the Voted and Not Voted group. The null hypothesis was that each mean implicit measure had no difference in the Voted and Not Voted groups. To test the hypothesis, we applied the Mann-Whitney-Wilcoxon Rank Sum Test (MWW Test) to compare the implicit measures between the two independent sample groups.

The study compared multiple implicit measures between the voted and not voted group. This evaluation involved multiple hypotheses, which can result in increased Type 1 errors in the data set. To correct the Type 1 errors, this study applied the Bonferroni correction. Each individual hypothesis was tested at an adjusted statistical significance level.

Logistic Regression Analysis

To clarify the relationship between the implicit measures and the patient's voting behavior, the study also conducted logistic regression analysis.

Page Staying Time served as the primary implicit measure for this study. We first built a regression model with Page Staying Time as the input variable and voting result as the outcome. This model controlled demographic variables including age, race and gender. Study participant's baseline hemoglobin A1c value was also included in the model. The voting outcome was set to 1 for the Voted group and 0 for the Not Voted group. Equation 1 shows this regression model in detail.

$$\ln\left[\frac{P(Y)}{1-P(Y)}\right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5$$

$P(Y)$	Probability of Getting Patient Vote
β_0	Coefficient of intercept
$\beta_1 - \beta_5$	Coefficient of the Input Variable
X_1	Page Staying Time (minutes)
X_2	Age (years)
X_3	Race (White, Black, Others)
X_4	Gender (Male, Female)
X_5	A1C level of the most recent measure before the study period

Equation 1 Logistic Regression with Page Staying Time as Input

In addition to Page Staying Time, this study also examined in the link based measures and their relationships with the voting behavior. We built a regression model based on four link measures including Sound Play, Drug Table, AHRQ Link, and Activity Link. The study excluded link measures with sparse data to simplify the model. The excluded link measures included Fast facts, AHRQ Share, Nutrition link, About Diabetes link and Messaging link. This model also controlled demographic variables including age, race and gender. Study participant's baseline hemoglobin A1c value was included in the model. Equation 2 shows the details of the link measure based regression model.

$$\ln\left[\frac{P(Y)}{1-P(Y)}\right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8$$

$P(Y)$	Probability of Getting Patient Vote
β_0	Coefficient of intercept
$\beta_1 - \beta_8$	Coefficient of the Input Variable
X_1	Sound Play Measure
X_2	Drug Table Measure
X_3	AHRQ Link
X_4	Activity Link
X_5	Age (years)
X_6	Race (White, Black, Others)
X_7	Gender (Male, Female)
X_8	A1C level of the most recent measure before the study period

Equation 2 Logistic Regression with Link based Measures

CHAPTER IV

RESULTS

This study was conducted over a six-month period from June 22, 2012 to December 22, 2012. The existing iADAPT study population, a total of 1,450 patients, participated in the study during this six-month period. A total of 478 (33%) patients made one or more visits to the Diabetes Education page during the study period. There were 972 (67%) patients who did not visit the page. Table 4 shows demographic data for the study participants. There were not any significant demographic differences between patients who did and did not visit the Diabetes Education page.

Table 4 Study Population Demographic Summary - 1

Characteristic	Study Population (N=1450)		
	Visited the Page (N=478)	Did Not Visit the Page (N=972)	P value
	N (%) or Mean (SD)	N (%) or Mean (SD)	
Age (years)	59.87(11.28)	60.64 (12.39)	0.15
Gender			
Female	238 (49.79)	484 (49.79)	0.99
Male	240 (50.21)	488 (50.21)	
Race			
White	371 (77.61)	752 (77.37)	0.74
Black	86 (17.99)	160 (16.46)	
Others	21 (4.40)	60 (6.17)	
AlC	7.19 (1.46)	7.21 (1.82)	0.79

** *Statistically Significant is claimed when P-value is <0.05*
A1C is Hemoglobin A1C
SD is standard deviation

Among the patients who visit the MHaV Diabetes Education page, 120 (25%) patients responded to the survey by casting one or more votes while 358 (75%) patients visited the page but did not cast any votes. The study compared the demographic data of the Voted and Not Voted groups (Table 5). The data show that the average age for the Voted group is significantly older than the Not Voted group.

Table 5 Study Population Demographic Summary - 2

Characteristic	Participated (N=478)		
	Voted (N=120)	Not Voted (N=358)	P value
	N (%) or Mean (SD)	N (%) or Mean (SD)	
Age (years)	61.92 (11.63)	59.18 (11.09)	0.007**
Gender			
Female	52 (43.33)	186 (51.95)	0.13
Male	68 (56.67)	172 (48.05)	
Race			
White	90 (75.00)	281 (78.49)	0.74
Black	26 (21.67)	60 (16.76)	
Others	4 (3.33)	17 (4.75)	
AlC	6.97 (1.20)	7.27 (1.53)	0.07

** *Statistically Significant is claimed when P-value is <0.05*
A1C is Hemoglobin A1C
SD is standard deviation

Among the 972 non-visitors of the site, 238 patients never logged in to MHaV during the study period, and 734 patients had logged in to MHaV but had not viewed the Diabetes Education page. The distribution of the study population is illustrated in Figure 11.

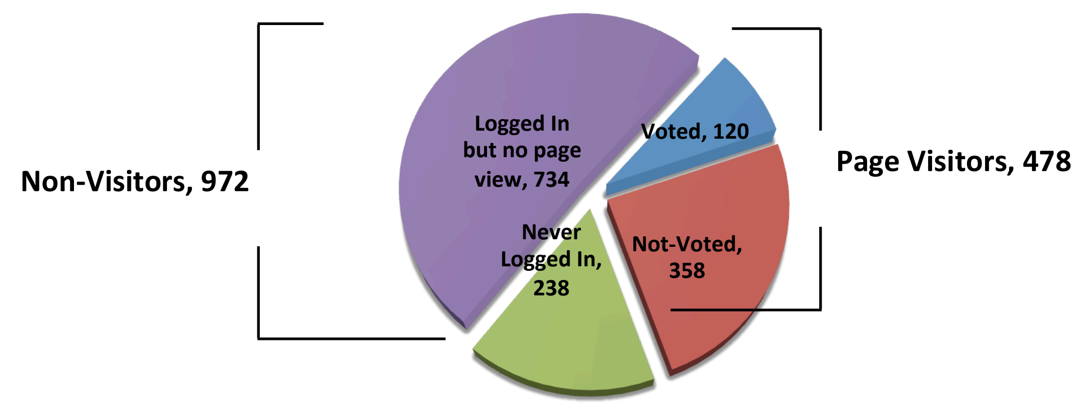

Figure 11 Distribution of Study Population in Pie Chart

Voting Outcome

Figure 12 illustrates the voting outcome from those study participants who visited the Diabetes Education page in a histogram. The data shows that 358 patients cast 0 votes, 57 patients cast 1 vote, 21 patients cast 2 votes, 32 patients cast 3 votes, 3 patients cast 4 votes, 4 patients cast 5 votes, 2 patients cast 6 votes, and 1 patient cast 7 votes.

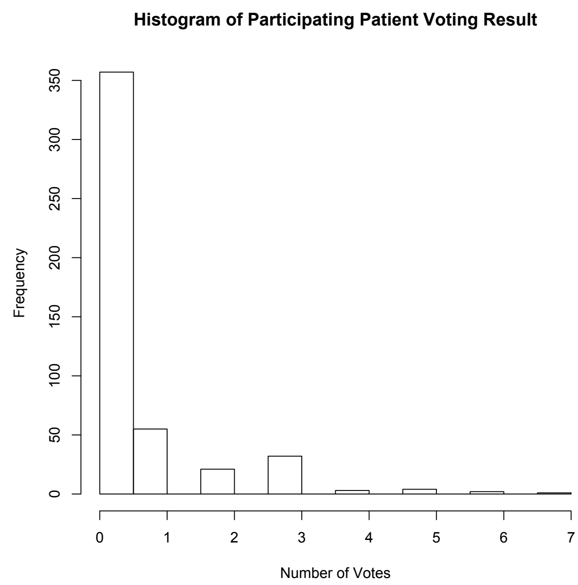

Figure 12 Distribution of the Voting Result in Histogram

Page Staying Time

Figure 13 illustrates the box plot of the distribution of the Page Staying Time in the Voted and the Not Voted groups. On average, the Not Voted group spent 3.66 minutes on the Diabetes Education page. The Voted group spent an average of 7.41 minutes on the page.

Figure 14 illustrates the histogram of the Page Staying Time in the Voted and the Not Voted groups. The histogram shows that the Page Staying Time distribution is highly skewed with long tails.

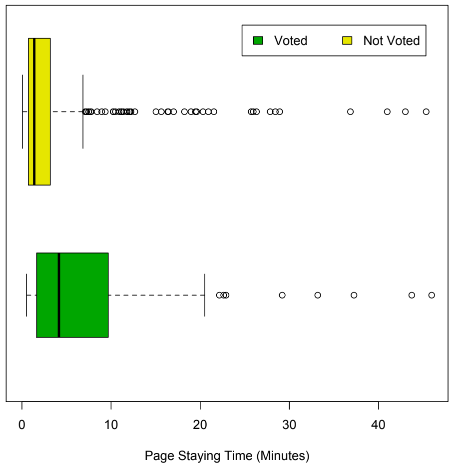

Figure 13 Distribution of the Page Staying Time as a Box Plot

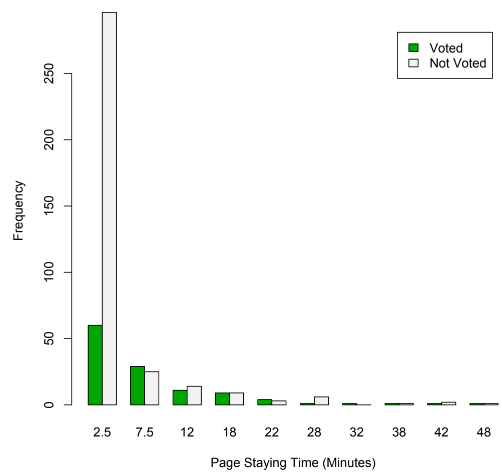

Figure 14 Distribution of the Page Staying Time as a Histogram

Link-based Measure

Figure 15 illustrates the box plot of the distribution of the Sound Play measure in the Voted and the Not Voted groups. On average, the Not Voted group members clicked the Sound Play button 0.51 times and the Voted group members clicked the sound play button 1.19 times.

Figure 16 illustrates the histogram of the Sound Play measure in the Voted and the Not Voted groups. The histogram shows the sound play measure distribution is highly skewed with long tails.

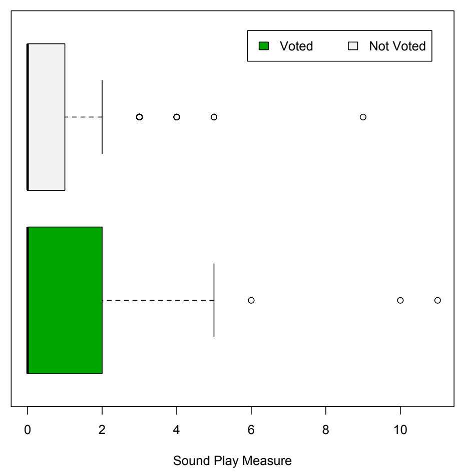

Figure 15 Distribution of the Sound Play Measure as a Box Plot

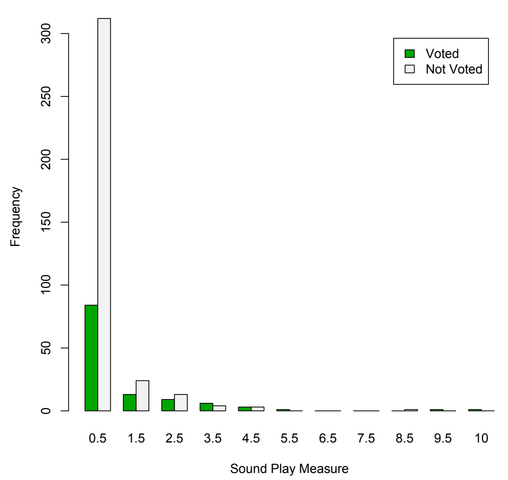

Figure 16 Distribution of the Sound Play Measure as a Histogram

The diabetes drug side effect comparison table was opened 0.3 times on average in the Voted group versus 0.1 times in the Not Voted group. The Voted group selected the AHRQ literatures links 1.2 times on average, compared to 0.4 times in the Not Voted group. The physical activity–related articles attracted an average of 0.4 views in the Voted group and 0.1 views in the Not Voted group.

Although varying in values, the link-based measures shared similar distribution traits to the Sound Play measure. The common traits can be summarized as 1) An asymmetrical distribution, 2) Highly skewed with long tails, and 3) Different variance between the Voted and the Not Voted groups.

Nonparametric Analysis Results

Table 6 displays the nonparametric analysis results from the Mann-Whitney-Wilcoxon Test. The results show that after the Bonferroni correction, there are five measures that differ significantly between the Voted and not Voted groups. These are: Page Staying Time, Sound Play measure, Drug Table measure, AHRQ link measure, and the Activity measure.

Table 6 Implicit Measure Summary and MWW Test Results

Independent Variable	Voted Group Median/Mean (1^{st} Qu., 3^{rd} Qu.)	Not Voted Group Median/Mean (1^{st} Qu., 3^{rd} Qu.)	P value *
Page Staying Time (Minutes)	4.15/ 7.41 (1.67, 9.64)	1.38/3.66 (0.73, 3.18)	<0.005**
Sound Play Measure	0.00/1.19 (0.00, 2.00)	0.00/0.51 (0.00, 1.00)	<0.005**
Drug Table Measure	0.00/0.36 (0.00, 1.00)	0.00/0.10 (0.00, 0.00)	<0.005**
AHRQ Link Measure	0.00/1.27 (0.00, 2.00)	0.00/0.41 (0.00, 0.00)	<0.005**
Activity Link Measure	0.00/0.49 (0.00, 0.00)	0.00/0.18 (0.00, 0.00)	<0.005**
Fast Facts Measure	0.00/0.24 (0.00, 0.00)	0.00/0.07 (0.00, 0.00)	0.013
Messaging Measure	0.00/0.08 (0.00, 0.00)	0.00/0.03 (0.00, 0.00)	0.018
About Diabetes Link Measure	0.00/0.07 (0.00, 0.00)	0.00/0.01 (0.00, 0.00)	0.096
Nutrition Link Measure	0.00/0.04 (0.00, 0.00)	0.00/0.03 (0.00, 0.00)	0.340
AHRQ Share Measure	0.00/0.03 (0.00, 0.00)	0.00/0.03 (0.00, 0.00)	0.411

* *Bonferroni-corrected P value with n = 10*
** *Statistically Significant is claimed when P-value is <0.005*

Logistic Regression Analysis Results

Table 7 shows the results of the logistic regression model with the Page Staying Time as the predictor and controlling for demographic data including age, race, gender, and patients' baseline hemoglobin A1C value. The results indicate that Page Staying Time has a significant positive relationship with the voting outcome. With each additional minute that a patient stays on the Diabetes Education page, the likelihood for the patient to vote increases 5.2%.

The results also indicate that patient's hemoglobin A1C value has a significant negative relationship with the voting outcome. Patients with a higher A1C value are less likely to vote. The likelihood of voting decreases 16% with one unit increase of the A1C value. The study calculated the overall predicting power of this regression model by measuring the area under the ROC curve (AUC). The AUC value for this time-based model is 0.66.

Table 7 Logistic Regression Result for Page Staying Time

Input Variable	Odds Ratio (OR)	95% Confidence Interval (CI)	P value
Page Staying Time (Minutes)	1.052	(1.025, 1.081)	0.000**
Age (years)	1.015	(0.995, 1.036)	0.146
Race=White	0.627	(0.356, 1.120)	0.108
Race=Others	0.597	(0.154, 1.898)	0.411
Gender=Male	1.416	(0.904, 2.231)	0.131
A1C	0.837	(0.703, 0.985)	0.038**

** *Statistically Significant is claimed when P-value is <0.05*

Table 8 shows the results of the logistic regression model with four link-based measures as predictors and controlling for demographic data including age, race, gender, and the baseline A1C value. The study found three link measures have significant positive relationship with the voting outcome: 1) Drug Table - the likelihood to vote increased 84% by with an increase in the click count of the drug comparison table by 1; 2) AHRQ literature link - with each additional click on the AHRQ literature link, the likelihood for patient to vote increased 30%; and 3) Activity Link - with each additional click on the Activity related reading links, the likelihood to vote increased 38%.

We also observed that the patient's hemoglobin A1C value had a significant negative relationship with the voting outcome. Patients with higher A1C value were less likely to vote. The likelihood to vote decreased 18% with each unit increase in the A1C value. The study calculated the overall predicting power of this regression model by measuring the area under the ROC curve (AUC). The AUC value for the link-based model is 0.69.

Table 8 Logistic Regression Result with Link-based Measures

Input Variable	Odds Ratio (OR)	95% Confidence Interval (CI)	P value
Sound Play	1.104	(0.921, 1.325)	0.284
Drug Table	1.844	(1.141, 3.016)	0.013**
AHRQ Link	1.304	(1.092, 1.570)	0.004**
Activity Link	1.380	(1.047, 1.858)	0.026**
Age (years)	1.016	(0.994, 1.038)	0.140
Race=White	0.561	(0.312, 1.024)	0.056
Race=Others	0.519	(0.123, 1.783)	0.328
Gender=Male	1.356	(0.850, 2.173)	0.203
A1C	0.824	(0.687, 0.977)	0.031**

** *Statistically Significant is claimed when P-value is <0.05*

CHAPTER V

DISCUSSION

This thesis proposes a new approach for measuring users' interest in web-based health literature by observing interactions with a web page through measures of implicit feedback. This section summarizes the key findings from an evaluation of this approach, the informatics contributions, the limitations of the work, and future directions, as well as the potential to expand the research from patient portal to public web sites.

Key Findings

Evaluating the relationship between a user's interests in web-based health literature and the amount of time the patient stays on the web page is the main focus of this research project. In a study of patients' use of a Diabetes Education web page within a patient portal, we observed that the amount of time a patient stayed on the page was significantly different between patients who vote and do not vote on the usefulness of the page. The patients who voted on average spent twice the amount of time on the page compared to the group that did not vote. The Page Staying time had a significant positive relationship with the patient's voting behavior. Positive voting behavior is known to be associated with user interest, and thus, the implicit feedback of Page Staying Time can be used as a measure of user interest. [64,65]

The study also found that link-based measures were significantly different between the voting and the not voting groups. Patients are more likely to vote when they had a higher number of activities with the sound play button, the AHRQ literature, and the activity reading links.

Informatics Contribution

The work presented in this thesis makes several useful informatics contributions in the area of measuring patients' interest in web-based health information.

The study of implicit feedback and the study of patients' interests are not new, but the concept of applying implicit feedback in evaluating patients' interests has not been explored by previous research. This study appears to be one of the first efforts in applying implicit feedback in evaluating patients' interests in web-based literature.

The second contribution is the development of a tracking tool that automatically collects, processes, and persists implicit feedback events without interrupting the users' interactions with the web page. In this study, the tracking tool measured two specific types of implicit feedback: Page Staying Time and Link Count. Although these two types of feedback are the first to be examined by this study, the tracking tool is designed with ability to handle a wide range of feedback events. The application of this tracking tool can also be extended to any web pages that support HTML and JavaScript.

The third contribution of this work is the development of a partial-time accumulation algorithm to handle the time calculation on abandoned web pages. Compared to existing algorithms as adapted by Google Analytics, this partial-time algorithm tracks every user-page event and accounts for these events in the time calculation. As a result, this new algorithm is able to provide time calculation on abandoned pages. This algorithm can also be expanded to work with generic web pages.

Limitations

This work has several limitations. First, the study limited the types of implicit feedback captured in the user's browser. In an ideal world, the users' interactions with web pages would be captured at the client side through a customized web browser or browser plugin. A customized web browser or browser plugin can provide the most comprehensive and accurate user activity data. For instance, a user's movement of the mouse, use of scroll bars, keyboard activity, and whether a mouse movement is inside or outside of a browser window would all be tracked. In reality, enforcing users to adopt a specialized browser or download a browser plugin is not practical and negatively affects the user's natural browsing behavior. This study limited the types of implicit feedback to Page Staying Time and content specific link measures, which can be measured in most browsers.

Second, there are several uncontrolled factors that could potentially affect results of this study. Regression models were built with controlled variables including age, race, gender and the baseline A1C value. However this study did not include the patient-specific factors such as Internet speed, computer skills, and literacy level. These factors could have affected study outcome.

Third, a user's interest is inherently a continuous variable. To simplify the analysis, this study flattens the interest measure into a Boolean outcome. This remains one limitation of this study.

Finally, a large portion of the patient population did not vote in the study. The reference standard for patient interest was based on user voting behaviors. Thus, the results could potentially be affected by a non-response bias. Further, we used voting as a global measure of interest in the page, regardless of the outcome of that vote about an individual component on the page. The utility of the implicit feedback to assess interest in individual parts of the page requires further investigation. To assess the potential non-response bias, we plan to do a follow-up study for users who did not vote. An electronic survey questionnaire will be sent to all of the study participants who did not vote. The survey would inquire about the patients' perception of the usefulness of the individual diabetes literature resources hosted at MHaV. We could then associate the survey results with their corresponding implicit feedback. These data would help the research team to have a better understanding of the nonresponse bias as well as user interest across the complete study population.

Future Directions

This research project evaluated Page Staying Time and Link Count. Another implicit measure that would be interesting to study in the future is the Link Ratio. This value measures the total number of user-clicked links normalized by the total number of links on a web page. A higher link ratio indicates that more links on a web page had been selected during a user visit. A high amount of user-web activity could potentially infer user interest.

Confined by time and scope, this research project leveraged an existing Diabetes Education page that was deployed in a patient portal and aimed towards patients with

Type II diabetes. The research project selected the Page Staying Time as the main implicit feedback metric because time is a generic measure that can be expanded and measured on most web pages. The tracking tool and partial time algorithm in this study are also developed to work with any web sites that support HTML and JavaScript. In the future, these metrics could be evaluated in a broad range of health consumers and general Internet users for a wide variety of health information resources.

CHAPTER VI

CONCLUSION

This thesis reports a study that evaluated the feasibility of measuring patients' interest in web-based health literature by tracking implicit feedback. The concept of implicit feedback is not new. Research in the Information Retrieval and Search Engine Optimization domains has found that the use of implicit feedback is an effective approach in evaluating users' opinions and preferences. [74] The study of patients' interest is not a new field either. User surveys and cognitive studies have been employed to measure patients' interest. This research innovates by measuring implicit feedback in an unexplored domain: evaluating patients' interest in web-based health literature. We hypothesized that a patient's interaction with a web page, reflected in terms of implicit feedback, was correlated with the patient's interest in the contents of the web page.

The study selected two types of implicit measures. The first measure was the total amount of time a patient stays on the web page. The second measure was the number of times a patient accesses a particular link on the web page. The patient's interest was measured as a binary variable: interested and not interested, as determined by whether a patient voted on a user survey, regardless of the nature of the vote.

Mann-Whitney-Wilcoxon nonparametric tests were done on both types of implicit measures. The results showed that both the time-based and the link-based measures differed significantly between the interested and not interested groups. Patients who were interested in the study page spent twice the amount of time compared to the patients who were not interested. Also, patients who were interested in the study page had a higher number of link counts compared to the patients who were not interested.

To further understand the relationship between the implicit measure and patients' interest, the study constructed two logistic regression models, one with each type of implicit measure as a predictor. Page Staying Time had a significant positive relationship with patients' interest. When a patient spends one extra minute on the study page, the likelihood for the patient to show interest increased 5.2%.

The three link-based measures had a significant positive relationship with patients' interest. The drug comparison table link is a strong predictor of patients' interest. The likelihood for a patient to show interest increased 84% when the click count of the drug comparison table increased by 1. The likelihood for a patient to show interest increased 30% when the click count of the AHRQ literature link increases by 1. The likelihood for a patient to show interest increased 38% when the click count of the physical activity reading link increases by 1.

The study discovered that patients' age, gender, and rate did not appear to have a significant relationship with patients' interest. However, patients' baseline hemoglobin A1C value has a significant negative relationship with interest, as measured by voting. Patients with a higher A1C value showed less interest in the study page.

In conclusion, the study found that patients' interest in the study page is associated with the amount of time patients spend on the page and the number of link-counts performed on the study page.

To summarize, the primary advantage of the implicit feedback approach is that it removes the cost to the patient of providing explicit feedback. The challenge and tradeoff lies in its accuracy compared to the explicit measures as in the case of user surveys and cognitive studies. Because large quantities of implicit data can be gathered at no extra cost, our study considers implicit feedback to be a promising alternative in evaluating patients' interest in web-based health literature.

REFERENCES

1. Health Online 2013. Fox, S., Duggan M. (2013). Retrieved from http://pewinternet.org/Reports/2013/Health-online/Summary-of-Findings.aspx

2. The Decade of Health Information Technology: Delivering Consumer-centric and Information-rich Health Care (2004). Thompson G. T., Brailer D. J. Retrieved from http://www.providersedge.com/ehdocs/ehr_articles/The_Decade_of_HIT-Delivering_Customer-centric_and_Info-rich_HC.pdf

3. Consumer health information seeking on the Internet: the state of the art. Cline, R. J. W., Haynes K. M. (2001) Retrieved from http://her.oxfordjournals.org/content/16/6/671.full

4. A Model for the Design and Implementation of a Participant Recruitment Registry for Clinical Studies of Older Adults. Dowling, N. M., Olson, N., Mish, T., Kaprakattu, P., Gleason, C. (2012) Retrieved from http://www-ncbi-nlm-nih-gov.proxy.library.vanderbilt.edu/pmc/articles/PMC3325341/

5. Innovations in Survey Research – an Application of Web-Based Surveys. Sills, S. J., Song, C. Social Science Computer Review. Spring 2002 Vol. 20 No. 1 pp. 22-30

6. Working with Low Survey Response Rates: The efficacy of Weighting Adjustments, Dey. E.L., Research in Higher Education, April 1997, Volume 38, Issue 2, pp: 215-227

7. eHealth Literacy: Essential Skills for Consumer Health in a Networked World, Norman, C.D., Skinner, H.A. J Med Internet Res. 2006 Apr-Jun; 8(2): e9 doi: 10.2196/jmir.8.2.e9

8. A Holistic Framework to Improve the Uptake and Impact of eHealth Technologies, Gemert-Pijnen, J.E., Nijland, N., Limburg. M, Ossebaard, H.C., Kelders, S.M., Eysenbach. G., Seydel, E. J Med Internet Res. 2011 Oct-Dec; 13(4): e111 doi: 10.2196/jmir.1672

9. Using the Internet for surveys and health research, Eysenbach, G., Wyatt, J. J Med Internet Res. 2002, Apr-Nov; 4(2): E13

10. Working with Low Survey Response Rates: The efficacy of Weighting Adjustments, Dey. E.L., Research in Higher Education, April 1997, Volume 38, Issue 2, pp: 215-227

11. AAPOR Response Rate – An Overview. Retrieved from http://www.aapor.org/Response_Rates_An_Overview1.htm

12. Nonresponse Bias in Household Surveys. Singer, E. Public Opinion Quarterly, volume 70, Issue 5, 2006

13. Gauging the Impact of Growing Nonresponse on Estimates from a National RDD Telephone Survey, Keeter, S., Kennedy, C., Dimock, M., Best, J., Craighill, P. Public Opinion Quarterly, volume 70, Issue 5, 2006 pp: 759-779

14. Increasing Survey Response Rates: Combining Experimental Manipulations, Buyer, L.S., Miler, K.J. Retrieved from http://www.govst.edu/uploadedFiles/Institutional_Research/Survey%20Response%20Rates%206.pdf

15. Overcoming Survey Research Problems. Porter, S.R. 2004, New Directions for Institutional Research, Number 121, pp.103

16. A comparison of response rate, response time, and costs of mail and electronic surveys. Shannon, D.M., Bradshaw, C.C., 2002, Journal of Experimental Education, 70, pp: 179-192

17. Data That Do Not Persuade: Results of an Incentive Experiment, Fecso, R.S., Broach, R.J., Grigorian, K.H. National Science Foundation. Retrieved from http://www.fcsm.gov/05papers/Fecso_VIIB.pdf

18. Mail survey response rate: a meta-analysis of selected techniques for inducing response. Fox, R.J., M.R., Kim, J. Public Opinion Quarterly, vol.52, pp.467-491

19. Factors Affecting Response Rates to Mailed Questionnaires: A Comprehensive Literature Review. Jarvey, L. Journal of the Market Research Society, vol.29, pp.341-353

20. Mode and incentive effects on aspects of survey administration and data quality. O'Brien, J.E., Levin, K.Y., Hagerty-Heller, t., Greenlees, J.B., Kirsch, A.D. 2001. Retrieved from http://www.amstat.org/sections/srms/Proceedings/y2001/Proceed/00365.pdf

21. Nonresponse Rates and Nonresponse Bias in Household Surveys, Groves R.M., American Association for Public Opinion Research, doi: 10.1093/poq/nfl033

22. Usability in the real world: assessing medical information technologies in patients' homes, Kaufman D, Patel V, Hillman C, Morin P, et al, Journal of Biomedical Informatics, Volume 36, Issues 1-2, February-April 2003, pp: 45-60

23. Usability testing in medical informatics: cognitive approaches to evaluation of information systems and user interfaces, Kushniruk A.W, Patel V.L, Cimino J.J, Proc. AMIA Annu. Fall. Symp. (2001), pp. 163-173

24. eHEALS: The eHealth Literacy Scale, Norman, C, Skinner H, J Med Internet Res. 2006 Oct-Dec; 8(4):e27.

25. Accurately Interpreting Clickthrough Data as Implicit Feedback, Joachims T, Granka, L, Pan B, SIGIR'05, August pp:15-19

26. Gauging the Impact of Growing Nonresponse on Estimates from a National RDD Telephone Survey, Keeter, S., Kennedy, C., Dimock, M., Best, J., Craighill, P. Public Opinion Quarterly, volume 70, Issue 5, 2006 pp: 759-779

27. Implicit Feedback for Recommender Systems. In Proceedings of the AAAI Workshop on Recommender Systems, Oard, D., Kim, J., July 1998, pp. 81-83

28. Optimizing Search Engines using Clickthrough Data, Joachims, T, Technical report, Department of Computer Science, 2002

29. Measuring retrieval effectiveness based on user preference of documents, Yao,Y, Journal of the American Society for Information Science, 46(2):133-145, 1995

30. Unbiased evaluation of retrieval quality using clickthrough data, Joachims, T, Technical report, Department of Computer Science, 2002

31. Toward a more informed patient: bridging health care information through an interactive communication portal, Koonce T, Giuse D, Beauregard J, Giuse N, Journal of the Medical Library Association, v.95(1), Jan 2007, pp: 77-81

32. MyHealthAtVanderbilt: policies and procedures governing patient portal functionality, Osborn C, Rosenbloom S, Stenner S, Anders S, Muse S, Johnson K, Jackson, G, JAMIA, Volumn 18, Issue Suppl 1

33. Web Surveys A review of issues and approaches, Couper M, The Public Opinion Quarterly, Vol.64, No.4, Winter, 2000, pp: 464-494

34. HTML Introduction. Retrieved from http://www.w3schools.com/html/html_intro.asp

35. The Internet Protocol. Retrieved from http://www.rfc-editor.org/rfc/rfc791.txt

36. Introducing JSON. Retrieved from http://www.json.org/

37. The application/json Media Type for JavaScript Object Notation. Retrieved from http://www.ietf.org/rfc/rfc4627.txt?number=4627

38. JSON The Fat-Free Alternative to XML. Retrieved from http://www.json.org/fatfree.html

39. AJAX Introduction. Retrieved from http://www.w3schools.com/ajax/ajax_intro.asp

40. Mastering Ajax, Part 2: Make asynchronous requests with JavaScript and Ajax. Retrieved from http://www.ibm.com/developerworks/web/library/wa-ajaxintro2/?ca=dgr-lnxw07AJAX-Request

41. Oracle Database 10g Release 2: A Revolution in Database Technology. Retrieved from http://www.oracle.com/technetwork/issue-archive/2005/05-sep/bwp-overview-10gr2-060205-131882.pdf

42. Oracle Database Documentation Library. Retrieved from http://www.oracle.com/pls/db102/homepage

43. Evaluation of patient satisfaction with tailored online patient education information. Atack L, Luke R, Chien E. Comput Inform Nurs, 2008 Sept.-Oct.; 26(5): p. 258-64.

44. Quality of Life: A Closer Look at Measuring Patients' Well-Being, Frank J. Snoek, MA, PhD, Diabetes Spectrum, volume 13 Number, 2000, p. 24.

45. Using technology to deliver healthcare education to rural patients. McIlhenny CV, Guzic BL, Knee DR, Wendekier CM, Demuth BR, Roberts JB. Rural Remote Health. 2011 Oct;11(4):1798. Epub 2011, October 11.

46. Implicit Ratings and Filtering, Nichols, D. M. In Proceedings of the Fifth DELOS Workshop on Filtering and Collaborative Filtering, Nov., pp. 221–228.

47. Information Filtering Based on User Behavior Analysis and Best Match Retrieval. Morita, M., and Shinoda. Y. In: proceedings of the SIGIR Conference.

48. GroupLens: Applying Collaborative Filtering to Usenet News. Konstan, J., B. Miller, D. Maltz, J. Herlocker, L. Gordon, and J. Riedl. (1997). Communications of the ACM. Vol. 40 (3): pp. 77–87.

49. Implicit Feedback for Recommender Systems. Oard, D., and J. Kim. (1998). In Proceedings of the AAAI Workshop on Recommender Systems, July, pp. 81–83.

50. Implicit feedback for inferring user preference: A bibliography. D. Kelly and J. Teevan, In SIGIR Forum, 2003.

51. Response rate and measurement differences in mixed mode surveys. Dillman, D.A., Phelps, G., Tortora, R., Swift, K., Kohrell, & Berck, J. Retrieved from http://survey.sesrc.wsu.edu/dillman/papers/Mixed%20Mode%20ppr%20_with%20Gallup_%20POQ.pdf

52. Learning Users' Interests by Unobtrusively Observing Their Normal Behavior. Goecks, J., and J. Shavlik. (1999). In Proceedings of the IJCAI Workshop on Machine Learning for Information Filtering, pp. 129–132.

53. Inferring User Interest. Claypool, M., D. Brown, P. Le, and M. Waseda. IEEE Internet Computing, November-December, pp. 32–39.

54. Optimizing Search Engines using Clickthrough Data. Joachims, T. (2002a). In Proceedings of KDD 2004, pp. 133–142.

55. Evaluating Implicit Measures to Improve Web Search, Steve Fox, Kuldeep Karnawat, Mark Mydland, Susan Dumains, and Thomas White, ACM Transactions on Information systems, Volume 23 Issue 2, April 2005.

56. Learning User Interaction Models for Predicting Web Search Result Preferences E. Agichtein, Eric Brill, Susan Dumais, Robert Ragno. SIGIR '06 Prceedings of the 29th annual International ACM SIGIR conference on Research and development in information retrieval. ISBN: 1-59593-369-7.

57. Improving Web Search Ranking by Incorporating User Behavior, E. Agichtein, E. Brill, and S. Dumais, in Proceedings of the ACM Conference on Research and Development on Information Retrieval (SIGIR), 2006.

58. Learning users' interests by unobtrusively observing their normal behavior. J. Goecks and J. Shavlick. In Proceedings of the IJCAI Workshop on Machine Learning for Information Filtering. 1999.

59. Optimizing Search Engines Using Clickthrough Data, T. Joachims, in Proceedings of the ACM Conference on Knowledge Discovery and Datamining (SIGKDD), 2002

60. Accurately Interpreting Clickthrough Data as Implicit Feedback, T. Joachims, L. Granka, B. Pang, H. Hembrooke and G. Gay, in Proceedings of the ACM Conference on Research and Development on Information Retrieval (SIGIR), 2005.

61. Total Prevalence of Diabetes and Pre-Diabetes. Retrieved from http://www.diabetes.org/diabetes-statistics/prevalence.jsp.

62. Cascading Style Sheets [Internet]. Available from: http://www.w3.org/Style/CSS/Overview.en.html.

63. Google Analytics documentation. Available from http://support.google.com/analytics/bin/answer.py?hl=en&answer=1006924

64. The Role of Topic Interest in Survey Participation Decisions. Groves, Robert M. Presser, Stanley, Dipko, Sarah. Public Opinion Quarterly, Volume 68, Issue 1, pp. 2–31.

65. The Influence of Topic Interest and Interactive Probing on Response to Open-Ended Question in Web Surveys. Jennifer L. Holland, Leah Melani Christian. Social Science Computer Review, Volume 27, Issue 2, pp. 196–212.

66. The Missing Link: Bridging The Patient–Provider Health Information Gap, Paul C. Tang and David Lansky, Health Affairs, 24, no. 5 (2005): pp. 1290–1295.

67. Personal Health Record: New Opportunity for Patient Education, Kupchunas*,* William R. Orthopedic Nursing Issue: Volume 26(3), May/June 2007, pp. 185–191.

68. Diabetes Education: Integrating Theory, Cultural Considerations and individually Tailored content*, Chandra Osborn, Jeffrey D. Fisher.*

69. Patient web portals to improve diabetes outcomes: a Systematic review. Osborn CY, Mayberry LS, Mulvaney SA, Hess R.

70. Toward a more informed patient: bridging health care information through an interactive communication portal, Taneya Y. Koonce, Dario A. Giuse, Julie M. Beauregard, Nunzia Giuse. J Med Library Assoc. 2007 January; 95(1): pages 77-81.

71. A framework for clinical communication supporting Healthcare Delivery, Jim Jirjis, Jacob B. Weiss, Dario Giuse, and S. Trent Rosenbloom. AMIA Annual Symp Proc. 2005; 2005:3-379.

72. Health coaching via an Internet portal for primary care patients with chronic conditions: a randomized controlled trial. Leveille SG, Huang A, Tsai SB, Allen M, Med Care. 2009 Jan; 4791): p. 41-7.

73. Measuring the Impact of Patient Portals: What the Literature Tells Us. Seth Emont, White Mountain Research Associates, http://www.chcf.org/publications/2011/05/measuring-impact-patient-portals

74. Evaluation of Information Retrieval Systems: Approaches, Issues, and Methods. Harter, Stephen P., Hert, Carol A. Annual Review of Information Science and Technology (ARIST), v32 p. 3-94, 1997.

75. Patient access to an electronic health record with secure messaging: impact on primary care utilization. Zhou, Yi Y, Garrido, Terhilda, Chin, Homer L, Wiesenthal, Andrew M, Liang, Louise L, The American Journal of Managed Care, Volume 13, Issue 7, pp. 418–424.

76. AJAX [Internet]. Available from: http://www.w3schools.com/ajax/default.ASP

Appendix A

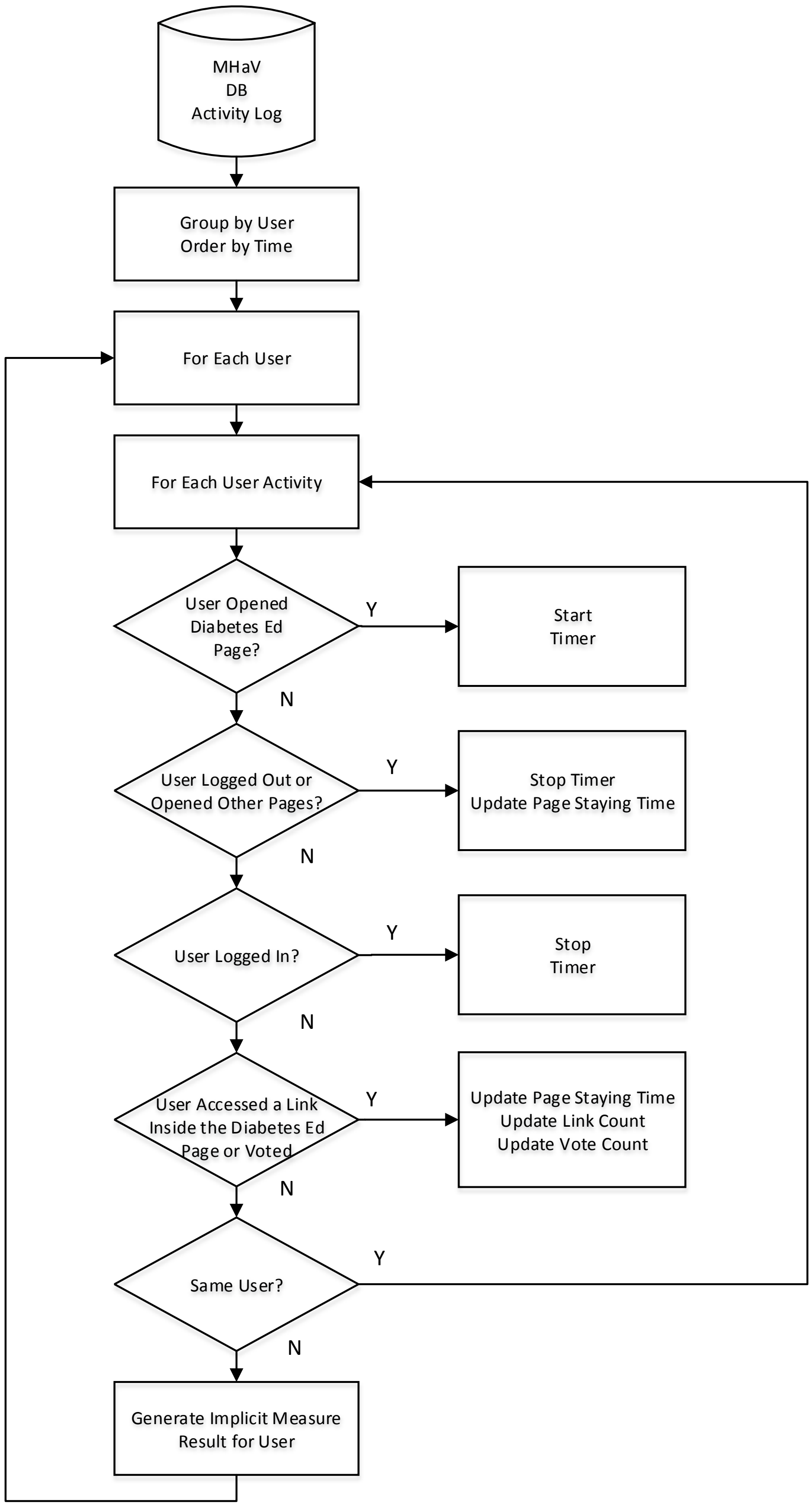

Figure 17 Calculation of Implicit Measure in Flow Chart